Variable Selection Results

Nirupama Tamvada

Simulation Settings: Simulating from Cause-Specific Cox Proportional Hazards Model

- We generate data from two cause-specific Weibull hazards according to Beyersmann et al. 2009
- The survival times for both causes are generated according to the following formula:

$$T = \left(- \frac{log(U)\lambda}{\exp(\beta X)} \right)^{1/v}$$

- Cause 1 is generated from a Weibull hazard with baseline hazard (λ) 0.55 (when all the covariates are set to 0) and the shape parameter (v) set to 1.5. Cause 2 is also generated from a Weibull hazard of baseline hazard 0.35, with the shape parameter also set to 1.5.
- The cause-specific indicator is generated from a binomial experiment with $p = \frac{\alpha_{01}}{\alpha_{01}+\alpha_{02}}$ (i.e the denominator represents the all-cause hazard)
- Any survival times less than 1/365 or greater than 1 were winsorized, i.e converted to be within a 1.5 year range (with white noise added so as to not have too many recurrent values)
- The X covariates are generated from a multi-variate normal distribution with $\mu = 0$ and a block correlation setting with 4 blocks with correlations 0.7, 0.4, 0.6, and 0.5. The noise covariates were generated with pairwise correlations 0.1.
- The censoring times are generated by an exponential distribution changed so that we have two censoring settings: 30 % censoring and 50% censoring

Models under comparison

- Cause-specific Independent Cox Model (with Elastic Net $\alpha = 0.7$) (enet-iCox)
- Cause-specific Cox with common penalty (with Elastic Net $\alpha = 0.7$) (enet-penCRCox)
- Casebase (with Elastic Net $\alpha = 0.7$) (enet-Casebase)
- Tapak et. al (2015) found elastic-net better than LASSO in a block-correlation setting

Simulation Parameters

1. N = 400
2. p = 120, p = 1000
3. Number of true covariates (Tp = 20)
4. Censoring: 30 % (leads to $\sim$ 47% cause of interest) and 50% (leads to $\sim$ 30% cause of interest)
5. Cause 2 > Cause 1

Simulation Settings

1. N = 400, p = 120, Varying effects between β_1 and β_2 censoring: 30 % , Cause 1 > Cause 2. This mimics biological biomarker data with varying block effects on both causes.
2. N = 400, p = 120, Varying effects between β_1 and β_2 censoring: 50 %, Cause 1 > Cause 2. This mimics biological biomarker data with varying block effects on both causes.
3. N = 400, p = 120, Varying effects between β_1 and β_2 censoring: 30 % , Cause 2 > Cause 1. This mimics biological biomarker data with varying block effects on both causes. Cause 2 is often greater in Cause 1 in sparse datasets where the risk curves cross.
4. N = 400, p = 1000, Varying effects between β_1 and β_2 censoring: 30 %, Cause 1 > Cause 2. This mimics biological biomarker data with varying block effects on both causes.
5. N = 400, p = 1000, Varying effects between β_1 and β_2 censoring: 50 %, Cause 1 > Cause 2. This mimics biological biomarker data with varying block effects on both causes.
6. N = 400, p = 1000, Varying effects between β_1 and β_2 censoring: 30 % , Cause 2 > Cause 1. This mimics biological biomarker data with varying block effects on both causes. Cause 2 is often greater in Cause 1 in sparse datasets where the risk curves cross.

7. N = 400, p = 120, effects only on cause 1 ($\beta_2 = 0$) censoring: 30 % , Cause 1 > Cause 2. This mimics the scenario where Cause 2 is death so there may not be any covariates associated with this.
8. N = 400, p = 120, opposite block effects of β_1 and β_2 censoring: 30 % , Cause 1 > Cause 2. This mimics the scenario of different biological pathways related to two relevant biological endpoints.
9. N = 400, p = 120, $\beta_1 = -\beta_2$ censoring: 30 % , Cause 1 > Cause 2. This scenario might resemble cancer therapy, i.e when patients are treated with chemo which may be associated with higher non-cause 1 mortality.
10. N = 400, p = 1000, effects only on cause 1 ($\beta_2 = 0$) censoring: 30 % , Cause 1 > Cause 2. This mimics the scenario where Cause 2 is death so there may not be any covariates associated with this.
11. N = 400, p = 1000, opposite block effects of β_1 and β_2 censoring: 30 % , Cause 1 > Cause 2. This mimics the scenario of different biological pathways related to two relevant biological endpoints.
12. N = 400, p = 1000, $\beta_1 = -\beta_2$ censoring: 30 % , Cause 1 > Cause 2. This scenario might resemble cancer therapy, i.e when patients are treated with chemo which may be associated with higher non-cause 1 mortality.

Simulation Settings: Misspecified Model (Proportional sub-distribution hazards = non-proportional cause-specific hazards)

- We generate data based on a two-cause model specification according to Fine and Gray (1999).
- Any survival times less than 1/365 or greater than 1 were winsorized, i.e converted to be within a one-year range (with white noise added so as to not have too many recurrent values)
- The causes were generated through a binomial experiment with $p = (1 - p_0^{e^{X\beta_1}})$ with p_0 set to 0.6 to generate Cause 1 as the cause with the largest incidence
- The censoring times are generated by a uniform distribution $U[0, M]$ with M changed so that we have two censoring settings: 30 % censoring and 50% censoring
- The X covariates are generated from a multi-variate normal distribution with $\mu = 0$ and a block correlation setting with 4 blocks with correlations 0.7, 0.4, 0.6, and 0.5. The noise covariates were generated with pairwise correlations 0.1.

Models under comparison

- Cause-specific Independent Cox Model (with Elastic Net $\alpha = 0.7$) (enet-iCox)

- Cause-specific Cox with common penalty (with Elastic Net $\alpha = 0.7$) (enet-penCRCox)
- Casebase (with Elastic Net $\alpha = 0.7$) (enet-Casebase)

Simulation Parameters

1. N = 400
2. p = 120, p = 1000
3. Number of true covariates (Tp = 20)
4. Censoring: 30 % (leads to $\sim$ 47% cause of interest)

Simulation Settings

13. N = 400, p = 120, Varying effects between $\beta_1 = -\beta_2$ censoring: 30 %
14. N = 400, p = 1000, Varying effects between $\beta_1 = -\beta_2$ censoring: 30 %